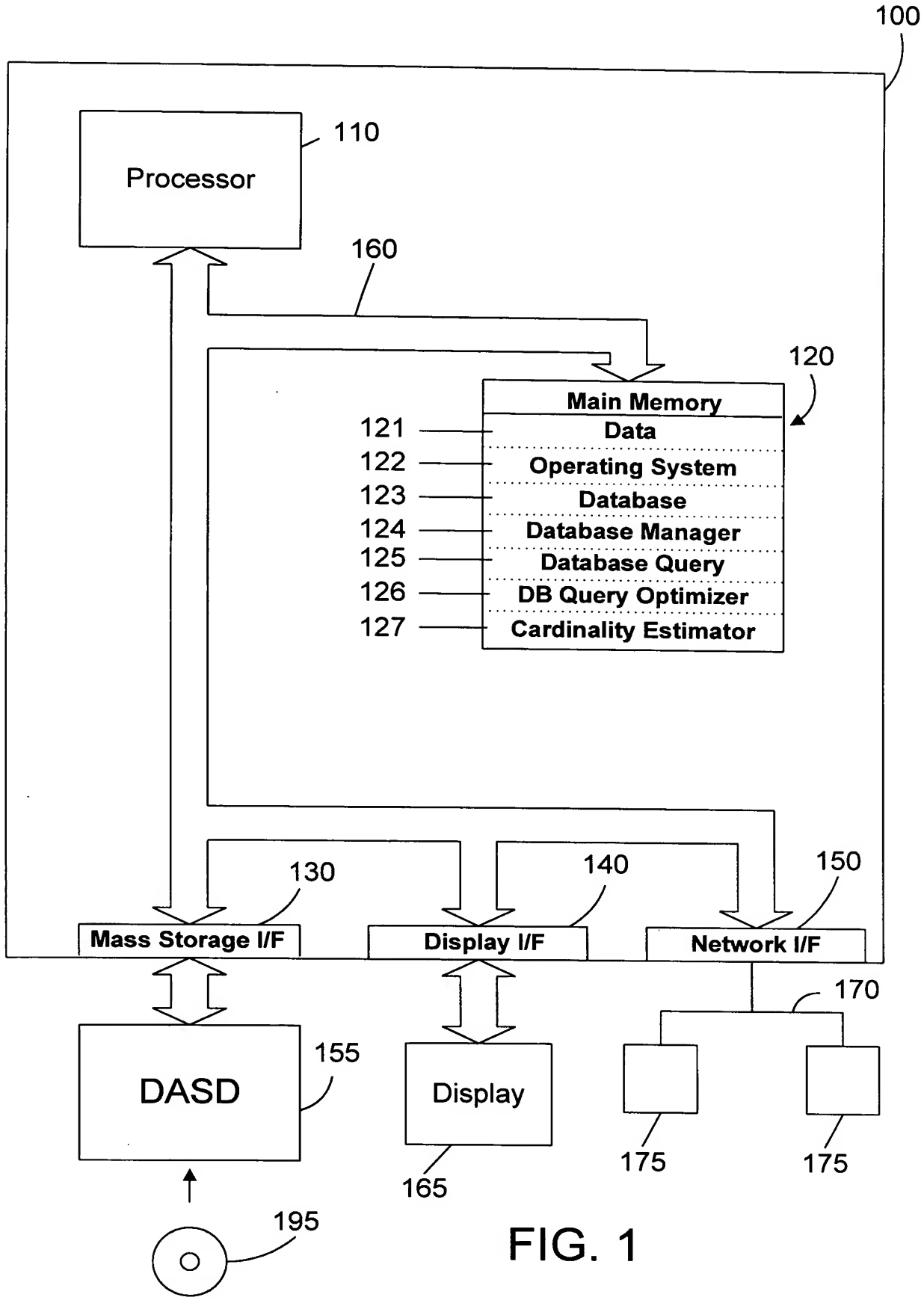

100
Processor
110
160
120
Main Memory
121
Data
122
Operating System
123
Database
124
Database Manager
125
Database Query
126
DB Query Optimizer
127
Cardinality Estimator
130
140
150
Mass Storage I/F
Display I/F
Network I/F
170
DASD
155
Display
165
175
175
195

FIG. 1

Select count(*) from T where T.A < b group by T.A

FIG. 2

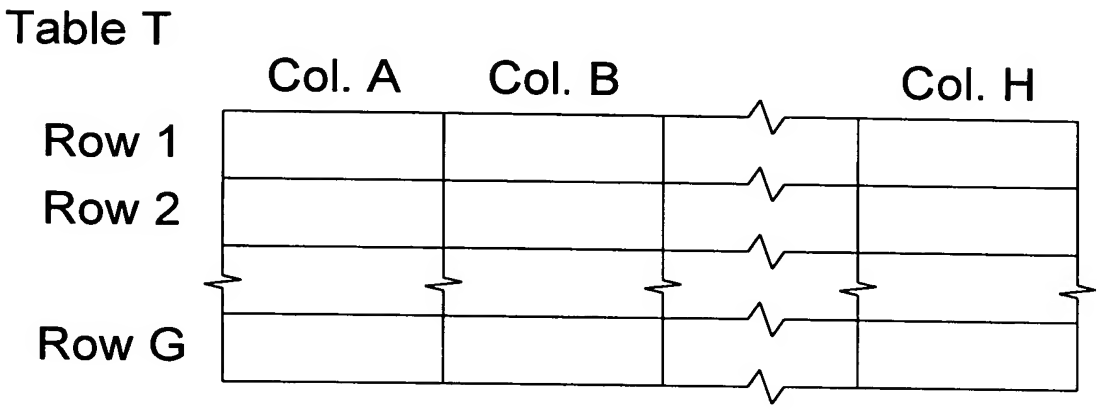


FIG. 3

Assumption: Cardinality of T.A is Ca

Query Optimizer Needs to Estimate Ca' (cardinality of T.A after applying local selection)

FIG. 4

$$Ca' = \min(Ca, X)$$

FIG. 5 Prior Art

$$Ca' = Ca(1-(1-1/Ca)^X)$$

where

X = Number of Rows in Intermediate Dataset

FIG. 6 Prior Art

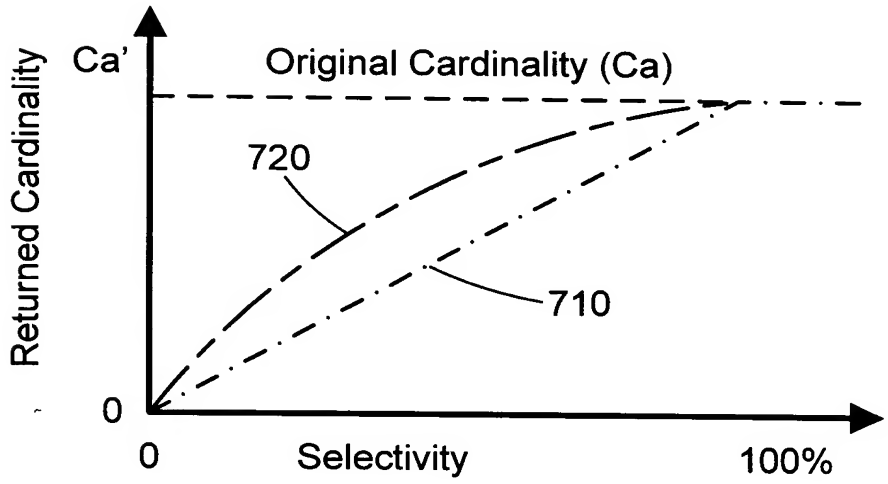


FIG. 7 Prior Art

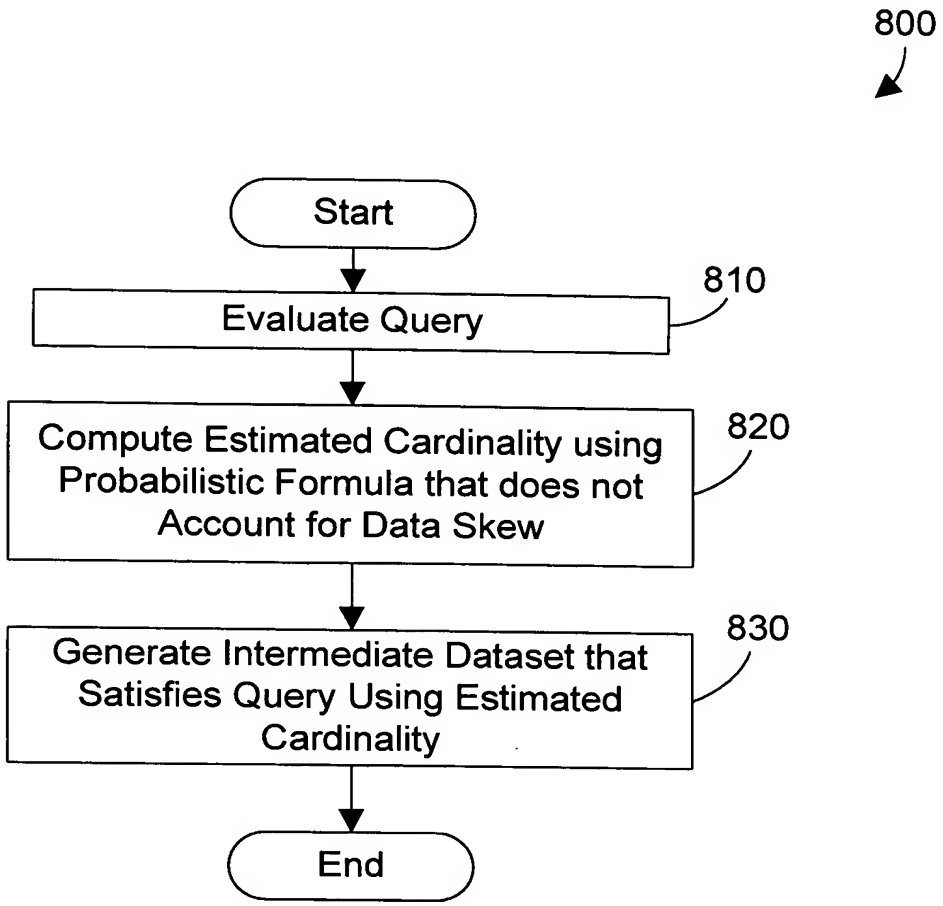


FIG. 8 Prior Art

$$Ca' = P + M(1-(1-1/M)^{Y})$$

where

$M = Ca - (P+Q)$

$Y = X - Fi$

X = number of rows in intermediate dataset

P = number of distinct skewed values in X

Q = number of distinct skewed values not in X

Fi = sum of frequencies for all skewed values in X above predetermined threshold

FIG. 9

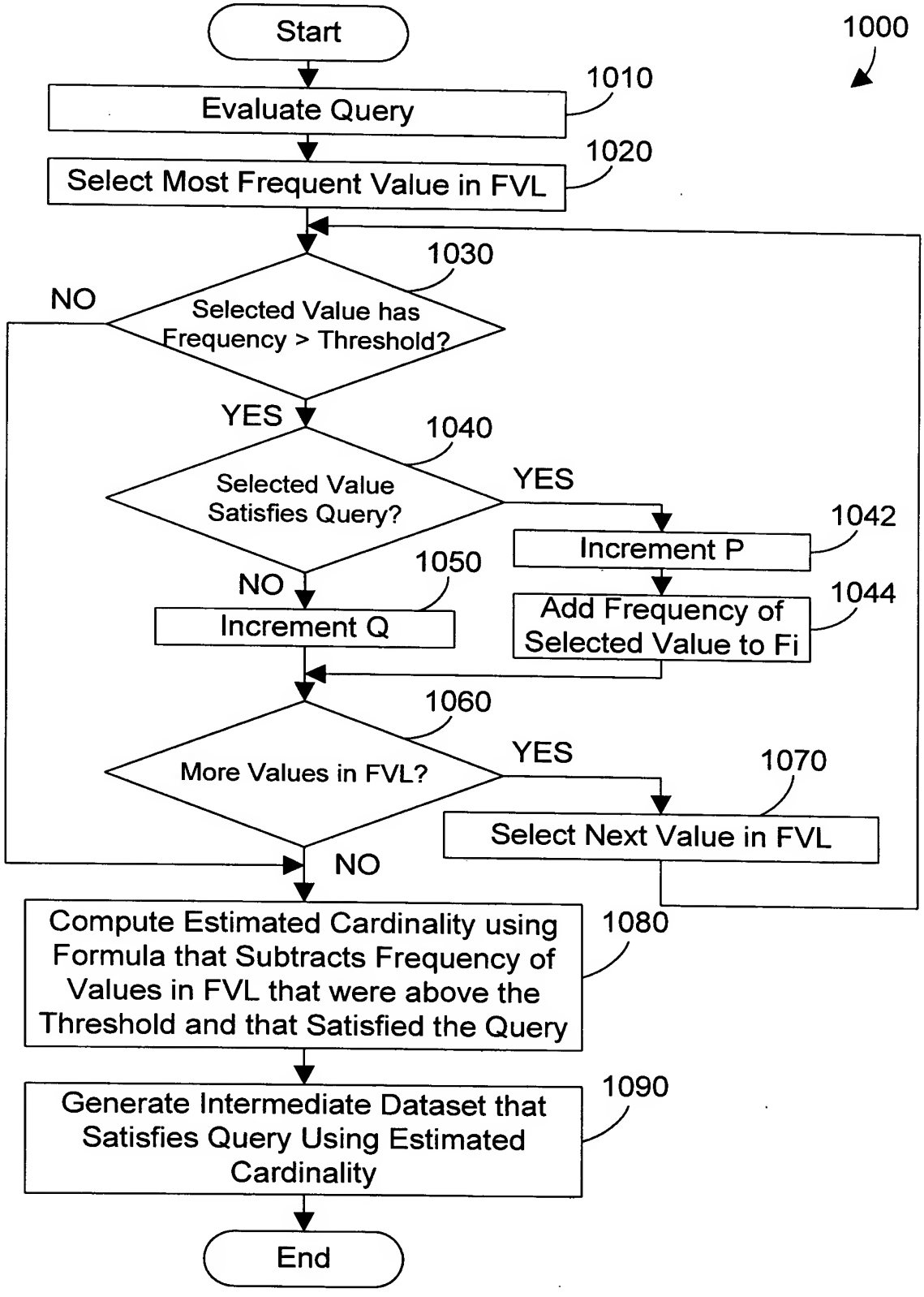

Start
1000
Evaluate Query
1010
Select Most Frequent Value in FVL
1020
1030
Selected Value has Frequency > Threshold?
NO
YES
1040
Selected Value Satisfies Query?
YES
1042
Increment P
1050
NO
Increment Q
Add Frequency of Selected Value to Fi
1044
1060
More Values in FVL?
YES
1070
Select Next Value in FVL
NO
Compute Estimated Cardinality using Formula that Subtracts Frequency of Values in FVL that were above the Threshold and that Satisfied the Query
1080
Generate Intermediate Dataset that Satisfies Query Using Estimated Cardinality
1090
End

FIG. 10

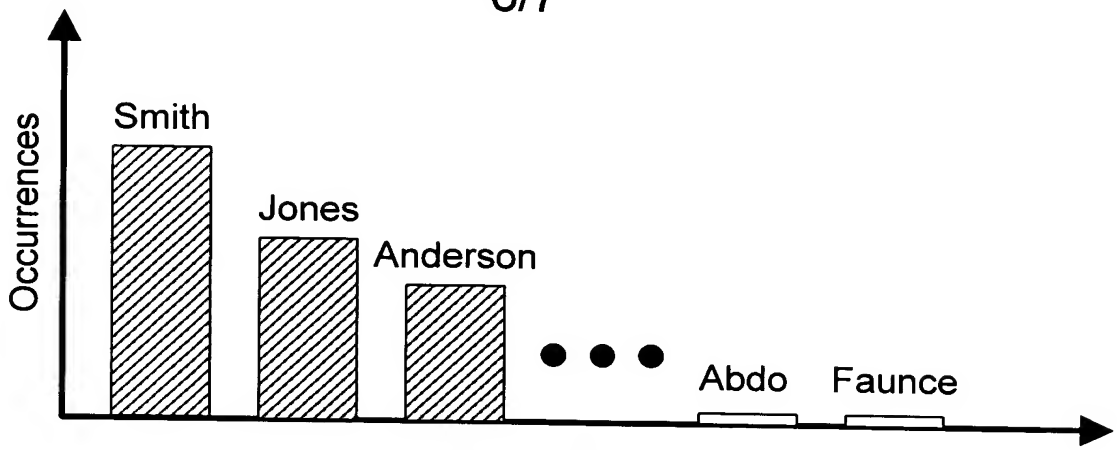
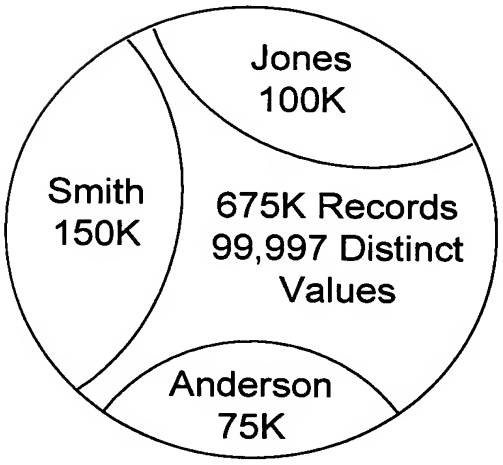


FIG. 11



Total Records = 1,000,000
Cardinality = 100,000

FIG. 12

Select count(*) from T where T.A > "F"

FIG. 13

Frequent Values List

1400

Value	Freq.
Smith	150,000
Jones	100,000
Anderson	75,000

1410

FIG. 14

$$Y = (1{,}000{,}000 * 0.333) - (150{,}000 + 100{,}000)$$
$$= 83{,}000$$

FIG. 15

$$M = 100{,}000 - (2+1)$$
$$= 99{,}997$$

FIG. 16

$$Ca' = 2 + 99{,}997(1-(1-1/99{,}997)^{83{,}000})$$
$$= 56{,}397$$

FIG. 17

$$Ca' = \min(100{,}000 , 333{,}000)$$
$$= 100{,}000$$

FIG. 18 Prior Art

$$Ca' = 100{,}000(1-(1-1/100{,}000)^{333{,}000})$$
$$= 96{,}421$$

FIG. 19 Prior Art